Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females

Rodrigo Gouveia-Oliveira, Anders Gorm Pedersen

www.elsevier.com/locate/yjtbi

Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females

Rodrigo Gouveia-Oliveira, Anders Gorm Pedersen

PII: S0022-5193(09)00270-7
DOI: doi:10.1016/j.jtbi.2009.06.028
Reference: YJTBI 5600

To appear in: *Journal of Theoretical Biology*

Received date: 24 September 2008
Revised date: 30 May 2009
Accepted date: 7 June 2009

Cite this article as: Rodrigo Gouveia-Oliveira and Anders Gorm Pedersen, Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2009.06.028

Higher variability in the number of sexual partners in males can contribute to a higher prevalence of sexually transmitted diseases in females

Rodrigo Gouveia-Oliveira, Anders Gorm Pedersen*

Center for Biological Sequence Analysis, Department of Systems Biology, The Technical University of Denmark, Building 208, 2800 Lyngby, Denmark

Abstract

By examining published, empirical data we show that men and women consistently differ in the shape of the distribution of the number of sexual partners. The female distribution is always relatively narrow - variance is low - with a big majority of women having a number of partners close to the average. The male distribution is much wider - variance is high - with many men having few sex partners and many others having more partners than most females.

Using stochastic modeling we demonstrate that this difference in variance is, in principle, sufficient to cause a difference in the gender prevalence of sexually transmitted diseases: compared to the situation where the genders have identical sex partner distributions, men will reach a lower equilibrium value, while women will stay at the same level (meaning that female prevalence becomes higher than male). We carefully analyse model behaviour and derive approximate expressions for equilibrium prevalences in the two different scenarios. We find that the size of the difference in gender prevalence depends on the variance ratio (the ratio between the variances of the male and female sex partner distributions), on the expected number of life-time partners, and on the probability of disease transmission. We note that in addition to humans, the variance phenomenon described here is likely to play a role for sexually transmitted diseases in other species also.

We also show, again by examining published, empirical data, that the

*Corresponding author

Email addresses: rodrigo@cbs.dtu.dk (Rodrigo Gouveia-Oliveira), gorm@cbs.dtu.dk (Anders Gorm Pedersen)

female to male prevalence ratio increases with the overall prevalence of a sexually transmitted disease (*i.e.*, the more widespread the disease, the more women are affected). We suggest that this pattern may be caused by the effect described above in highly prevalent sexually transmitted diseases, while its impact in low-prevalence epidemics is surpassed by the action of high-risk individuals (mostly males).

Key words:
Gender prevalence, STD, Epidemic, Variance, Agent-based modeling

1. Introduction

By definition, the total number of heterosexual contacts for men and women within a closed population has to be equal. If the population has exactly the same number of women and men, then the average number of sexual contacts for the two groups also has to be equal. However, there is no *a priori* reason why the distributions over the number of sexual contacts have to be identical for the two sexes. We were interested in whether any differences could be found between these distributions in real-world data, and in examining the influence of those potential differences in the gender prevalences of sexually transmitted diseases.

In Table 1 we present an overview of published, empirical data concerning the width of the distribution of the number of sex partners (Johnson et al., 1992, 2001; Melbye and Biggar, 1992; Lewin et al., 1998; Jæger et al., 2000; Haavio-Mannila et al., 2001). To the best of our knowledge, this represents the first overview of its kind and includes all publicly available data on the statistical dispersion (variance or interquartile range) of sex partner distributions from studies with unbiased sampling of subjects. (There are several additional studies where the number of sex partners is investigated but where only the average is reported.) It can be seen that all these empirical studies consistently show the same pattern: women are found to have relatively narrow (low variance) distributions, meaning most of them have close to the average number of sexual partners, while men have much wider (high variance) distributions. In a few studies the male sex partner distribution even appears to be bimodal (two-peaked; Jæger et al., 2000; Gregson et al., 2002).

The gender difference in the shape of the distribution of sexual partners is much clearer when one looks at the numbers for well-defined age-groups: since the number of sexual partners increases with age, merging data for

different age-groups can easily blur underlying patterns. One complication with comparing age groups is the empirically observed age mixing pattern: if women on average have sex with older men, then the distributions to compare could be those belonging to different age groups. Another complication is that the data presented in Table 1 results from self-reporting, and it is therefore likely to suffer from various biases (see Discussion). However, we can at the very least conclude that all available empirical data are consistent with there being a difference in the shape of sex partner distributions for men and women, thereby providing us with the motivation for further analysing the possible consequences of this phenomenon.

2. Monte Carlo Simulation

To investigate the effect of the differently shaped distributions on gender prevalence of sexually transmitted diseases, we performed Monte Carlo simulation of sexual interactions in heterosexual populations composed of equal numbers of women and men. Our model is agent-based (meaning every individual in the population is represented directly) and follows the spread of the disease over a series of discrete time-steps. The only difference between the two genders that is considered in this model is the distribution of the number of sexual contacts. In order to minimise the risk of programming artefacts, we constructed two entirely independent implementations of the model in the MATLAB and Python programming languages. The code is available from the authors upon request.

In our simulations, individuals interact once per time step. At every time step, each individual has a certain probability of establishing a sexual contact with an individual of the opposite sex. Each individual has its own probability of having a sexual contact, and this probability is fixed throughout the simulation. These probabilities may be different for different individuals and taken over the entire male or female population they follow a particular distribution. Specifically, we investigated scenarios using the following types of sex partner distributions: truncated normal, truncated mixture of two normals, Laplace, Gumbel, and discrete (with various number of categories). For each of these we experimented with a wide range of values for the expected number of partners per lifetime (N_p), the probability of disease transmission (p_{inf}), and the width (scale parameter) of the sex partner distributions. Whenever a sexual contact is established between an uninfected and an infected individual there is a certain risk of transmission. In our model,

this probability of transmission per sexual act (p_{inf}) is the same for all individuals. Individuals get older every time step until they reach a certain maximum age, when they die. A dead individual is immediately replaced by a new-born, sexually mature individual, that is not infected and that has no sexual experience. With respect to disease state, the model therefore has two compartments: S (susceptibles) and I (infectives). Once an individual is infected it remains so until it is removed from the population by death, at which point it is replaced by a new, susceptible, individual (note that both susceptible and infective individuals can die). The population size is constant and the age distribution uniform.

3. Results

3.1. Behaviour of model when genders have identical sex partner distributions.

The aim of this work is to investigate the effect of having differently shaped male and female sex partner distributions, but to gain a better understanding of the system we first performed a series of analyses where the male and female populations were identical in all respects. Among other things a number of simulations were performed where all individuals had exactly the same expected number of sex partners per life (*i.e.*, both genders had discrete sex partner distributions with only one category).

In this model, the basic reproduction number (the mean number of secondary cases caused by each infected individual in a naïve population) is given by: $R_0 = N_p p_{inf}$, where N_p is the expected number of partners per lifetime and p_{inf} is the probability that a contact between a susceptible and an infective leads to transmission. Consistently, simulations with a range of different p_{inf} and N_p values showed that the epidemic rapidly goes extinct when $N_p p_{inf} < 1$ (data not shown).

If simulation is allowed to continue for sufficiently many time-steps, the prevalence will eventually reach an equilibrium value (Figure 1). This equilibrium occurs when the removal of susceptibles by infection and the addition of new susceptibles by the birth-death process are balanced. When males and females have the same sex partner distributions, the equilibrium value does not depend on the shape of the distribution, but only on the values of p_{inf} and N_p. This was verified empirically by performing simulations with fixed p_{inf} and N_p, but with different sex partner distributions: discrete, normal,

mixture of two normals, and Gumbel distributions were tested, and all gave the same equilibrium prevalence.

To derive an expression for the size of the epidemic at equilibrium we first need to find an expression giving the change in the number of susceptible females (or males) during a single time-step:

$$\Delta S = -\frac{N}{2}\,(1 - f_I)\; f_{sex}\, f_I\, p_{inf} + \frac{N}{2}\,\frac{1}{N_{steps}}\, f_{dead,inf} \tag{1}$$

Here $\frac{N}{2}$ is the number of females in the population, f_I is the fraction of females (or males) that are infected, f_{sex} is the fraction of women (or men) having sex in a given time-step, N_{steps} is the number of time-steps per lifetime, and $f_{dead,inf}$ is the fraction of dead women that are infected.

The first part of equation 1 gives the decrease in susceptibles due to infection: There are $\frac{N}{2}\,(1 - f_I)$ susceptible women in all; multiplying by f_{sex} gives the number of susceptibles having sex in a given time-step; further multiplying by f_I gives the number of these that have sex with an infected male; and, finally, multiplying the result by p_{inf}, gives the number of cases where the disease was transmitted to the susceptible female. The fraction of women having sex in a time step is given by $f_{sex} = \frac{N_p}{N_{steps}}$.

The second part of equation 1 gives the increase in susceptibles due to replacement of dead, infected individuals with new, non-infected individuals: We are assuming a uniform age distribution, so each time-step the fraction $\frac{1}{N_{steps}}$ of the population is removed by death. The fraction $f_{dead,inf}$ of these are infected at the time of death. The exact value of $f_{dead,inf}$ is the same as the probability that an individual will have N_p partners and get infected, which is more easily computed as 1 minus the probability of having N_p partners *without* getting infected. Outside of the equilibrium state, this is somewhat complicated to determine since it relies on the set of different values of f_I at those time-steps where the individual has sexual contacts - essentially this is a binomial experiment with N_p trials but with constantly changing p. However, at equilibrium the value of f_I is constant by definition, and using that the probability of getting infected during a single sexual contact is $f_I p_{inf}$, we now find:

$$f_{dead,inf} = 1 - (1 - f_I p_{inf})^{N_p}$$

Inserting the expressions for f_{sex} and $f_{dead,inf}$ in equation 1 we get:

$$\Delta S = -\frac{N}{2}\,(1-f_I)\,\frac{N_p}{N_{steps}}\,f_I\,p_{inf} + \frac{N}{2}\,\frac{1}{N_{steps}}\,\left(1-(1-f_I p_{inf})^{N_p}\right)$$

At equilibrium the number of susceptibles is by definition constant, and we can therefore set this expression equal to zero and rearrange the equation, which gives us:

$$(1-f_I)\;N_p\,f_I\,p_{inf} = 1-(1-f_I p_{inf})^{N_p} \quad (2)$$

The value of interest, f_I, cannot easily be isolated in equation 2, but an approximate solution can be found by noting that, for reasonable values of N_p and p_{inf}, the right hand side of this equation is very close to 1. This approximation gives us:

$$-f_I^2 + f_I - \frac{1}{N_p p_{inf}} = 0 \quad (3)$$

From which the equilibrium frequency of infected individuals can finally be found:

$$f_I = \frac{1+\sqrt{1-4/\left(N_p p_{inf}\right)}}{2} \quad (4)$$

This approximate solution was found empirically to give very good predictions of the equilibrium prevalence when $N_p p_{inf} > 4$ (Figure 2; note that equation 3 has no real roots when $N_p p_{inf} < 4$).

3.2. Behaviour of model when genders have different sex partner distributions.

When males have sex partner distributions with higher variance than females, an interesting result is obtained: the gender equilibrium prevalences are no longer the same (Figure 3). Specifically, the male prevalence reaches a lower equilibrium value than the female prevalence. This phenomenon is reliably obtained with all investigated scenarios with higher variance for the male population (Figure 3, top three rows and data not shown. Also see Figure 4). However, a difference in prevalences is not obtained when the two genders have differently shaped distributions that have identical variances (Figure 3, bottom row).

In order to investigate the relationship between model parameters and the prevalence difference, a large number of simulation experiments were performed. In particular, we were interested in understanding how the difference between the male and female equilibrium prevalences depends on N_p, p_{inf}, and the variance ratio (the ratio between the variances of the male and female sex partner distributions). Figure 4 shows the equilibrium prevalence for males (coloured symbols) and females (black crosses) for different distribution types with a range of different values for N_p, p_{inf}, and the variance ratio. The dotted line on the plot indicates the equilibrium prevalence predicted using equation 4 (*i.e.*, the equation used to predict equilibrium prevalence when the two genders have identical sex partner distributions). All experiments depicted on Figure 4 had $N_p p_{inf} = 8$, and therefore have identical predicted prevalences. A number of conclusions can be made from Figure 4. First, the female prevalences (black crosses) can be seen to be identical to the predicted prevalence, regardless of simulation conditions. Secondly, variance ratios larger than one (*i.e.*, when the male sex partner distribution is wider than the female distribution) result in male equilibrium prevalences that are lower than the female prevalence. Third, for a given set of N_p and p_{inf} values, there seems to be a systematic relationship between the variance ratio and the male prevalence (and thereby the prevalence difference): note how similarly coloured data points (*i.e.*, data points from experiments with the same N_p and p_{inf} values) fall on a single line in the plot. Fourth, male prevalence apparently does not depend on the distribution type (and thereby on skew, kurtosis, and other higher moments of the distribution): note that for a given set of N_p and p_{inf} values (*i.e.*, for a given colour), experiments from all three distribution types fall on the same line.

Attempts to derive an analytical expression for the male prevalence from first principles proved to be very difficult (and the problem is perhaps not mathematically tractable). However, by carefully analysing the data in Figure 4 together with other similar data sets (with other values for N_p and p_{inf}) we managed to construct an expression that approximates the male prevalence fairly well:

$$f_{I,M} = f_{pred} - \frac{\mathrm{var}_{M/F}}{N_p^2 p_{inf}} + \frac{1}{N_p^{1.5}} - \frac{\mathrm{var}_{M/F}^2}{N_p^{2.5}} \tag{5}$$

Figure 5 shows the performance of equation 5 on a set of 60 data points covering three different distribution types, and a wide range of different values

of N_p, p_{inf} and the variance ratio. It can be seen that equation 5 captures the relationship quite well (Pearson $r^2 = 0.99$ for the 60 data points in Figure 5). While the overall performance of equation 5 is very good, it does become increasingly misbehaved as the three variables approach their lowest possible range: for instance the prevalence difference should be zero when the variance ratio is 1, but when both N_p and p_{inf} take on very low values, this is not the case for the above expression. For N_p and p_{inf} values slightly above their lowest range, however, equation 5 correctly predicts close to zero prevalence difference for a variance ratio of 1. This behaviour can also be seen on the plot above where the discrepancy between predicted and observed male prevalence increases with increasing prevalence difference. We conclude that equation 5 is a reasonable approximation for the model's behaviour for typical values of the three variables.

From the experiments summarised in Figure 3, 4, and 5 we conclude that a higher variance of the number of sex partners in males is sufficient to cause a difference in prevalence of sexually transmitted diseases between the two sexes. The effect is one where females have the same equilibrium prevalence as they would have if the two genders had identical sex partner distributions, while males end up with a lower equilibrium prevalence. From equation 5 (and from the empirical data it is based upon) we can see that the prevalence difference becomes larger if the variance ratio is increased, if N_p is decreased, and if p_{inf} is decreased.

In retrospect this effect is perhaps to be expected: the uneven distribution of male sex partner numbers must mean that a minority of men account for the majority of the total number of sexual encounters. These, more highly active, men have a high risk of becoming infected, and are in fact likely to have several encounters with infected women. Since they can only get infected once, the total rate of transmission from females to males therefore ends up being lower than if the two genders had identical sex partner distributions (because in that case more susceptible men would have become exposed to infected women). It should be stressed that while the phenomenon described in this paper has the apparent effect of giving females a higher prevalence, the actual effect is instead to lower the male prevalence (compared to the hypothetical situation where the two genders have identical sex partner distributions).

3.3. Empirical data on gender differences in prevalence of sexually transmitted diseases

The findings mentioned above prompted us to investigate empirical data on gender differences in prevalence or incidence of STDs to see if there are indications that the phenomenon described here plays a role in real-world epidemics. Listed in order of decreasing incidence, the 6 most common STDs in the U.S., as reported by the CDC, are: Human Papillomavirus (HPV), Herpes, Gonorrhoea, Hepatitis B, Syphilis and Chancroid (CDC, 2000). The 2 most incident of these ($>$ 1 million cases/year) have higher female incidence, the third (Gonorrhoea; $650,000$ cases/year) has very similar gender incidences, while the three least incident ($< 120,000$ cases/year) have higher male incidence (Nakashima et al., 1996; McQuillan et al., 1999; CDC, 2000). The highly incident Trichomoniasis and Chlamydia were excluded from this comparison, as the gender prevalence of the former is unknown while data from the latter is almost always biased by sampling. However, a study on Chlamydia in a non-clinical setting did observe women to have higher prevalence (Mertz et al., 1998), as also reported by, for instance, the CDC (CDC, 2000). It should be noted that these estimates of STD prevalence are associated with considerable uncertainty. However, the data do display an interesting pattern: female incidence is larger than male incidence in highly prevalent STDs, while the exact opposite is true in less prevalent diseases. The same can also be observed for HIV gender prevalences, when comparing countries with different prevalences. This relation between overall prevalence and gender prevalence has, again to the best of our knowledge, never been published before.

We suggest that the variance effect described above may play a role in the observed prevalence differences for highly prevalent diseases. There could be a number of different reasons for why the less prevalent diseases have higher incidence in men, including perhaps that these epidemics are driven mostly by the behaviour of high-risk groups (typically males). Under this hypothesis the variance effect only prevails after prevalence has reached a certain threshold and average members of the population start driving the epidemic.

4. Discussion

In this paper we have demonstrated three things: (1) All available empirical data on sex partner distributions are consistent with men having wider

(higher variance) sex partner distributions than women. (2) If men do in fact have a wider sex partner distribution than women, then this is in itself sufficient to cause a difference in the gender prevalence of sexually transmitted diseases (STDs): compared to the same-distribution scenario, men will reach a lower equilibrium prevalence, while women will stay at the same level. (3) Empirical data indicates that there is a relationship between how widespread an STD is, and its gender prevalence: the most incident STDs have higher prevalence in women, while the least incident STDs are more prevalent in men.

Regarding the empirical evidence for wider male sex partner distributions: as mentioned above these data are the result of self-reporting and are likely to suffer from a number of biases, including male over-reporting and female under-reporting (Wiederman, 1997; Alexander and Fisher, 2003; Nnko et al., 2004). It is also possible that the collected data do not adequately reflect sexual encounters resulting from prostitution. This could lead to an underestimation of female variance, although it should be pointed out that the relevant thing to measure is the amount of unprotected sex, which may be less frequent in the context of commercial sex, at least in developed countries (Ward et al., 1999, 2004; Wang et al., 2007). These are all relevant concerns, but we suggest that the important observation is that all available empirical data are consistent with male sex partner distributions being wider, and this in itself seems to be sufficient reason to further investigate the effects this condition would have on the spread of STDs.

Using a relatively simple agent-based, stochastic model we have shown that if men do have a wider sex partner distribution than women, then this in itself can lead to differences in the gender equilibrium prevalences. This has apparently not been noted or investigated before: While there exists an extensive literature on modeling STD epidemics (Anderson and May, 1988; Anderson and Garnett, 1996, 2000; Kretzschmar et al., 1996; Morris and Kretzschmar, 1997; Korenromp et al., 2000; Stover et al., 2002; Grassly and Garnett, 2005) only a few models account for differences in the shape or variance of the distribution of sexual contacts (Anderson and May, 1991; Garnett and Anderson, 1993, 1994; Thrall et al., 2000), and none of these analyses have explicitly commented upon the fact that the variance in males is higher than in females and investigated whether this could in itself be a causative factor for different gender prevalences. Perhaps part of the reason for this discrepancy has to do with modeling approach: the stochastic, agent-based approach employed here is directly aimed at situations where individual vari-

ability is important and makes it very easy to include age and sex partner structure in the population. However, this is much more cumbersome to account for in the compartmental epidemiological models (like the SIR model) that are often used, and where the focus is more on the behaviour of the average members of broad compartments. Finally, we would like to point out that the variance phenomenon described here is likely to play a role for sexually transmitted diseases in other species also, and that species-specific features such as unequal sex ratios may interfere with the effect in potentially interesting ways.

As mentioned above, we tentatively suggest that the variance effect may play a role in explaining the observed pattern where widespread STDs are more prevalent in women. We have also proposed that low-prevalence epidemics are mostly driven by high-risk individuals and their interactions, and that this could be the reason why these are more prevalent in men. While these hypotheses are consistent with the empirical data and with our other findings, there are of course several other plausible explanations for the phenomenon. Additionally, the data on gender prevalence are associated with considerable uncertainty. However, the evidence and hypotheses presented here at the very least indicates that further investigations could be of interest.

In conclusion, we suggest that the variance effect described in this paper should be explored further and perhaps included in models describing STD epidemics and in policy evaluations.

5. Acknowledgments

The authors gratefully acknowledge Professor Bo Lewin, Dept. of Sociology, University of Uppsala, Sweden, for kindly giving us access to original data from the 1996 Swedish sex survey.

References

Alexander, M. G., Fisher, T. D., 2003. Truth and consequences: using the bogus pipeline to examine sex differences in self-reported sexuality. J Sex Res 40(1), 27–35.

Anderson, R. M., Garnett, G. P., 1996. Low-efficacy HIV vaccines: potential for community-based intervention programmes. Lancet 348(9033), 1010–3.

Anderson, R. M., Garnett, G. P., 2000. Mathematical models of the transmission and control of sexually transmitted diseases. Sex Transm Dis 27(10), 636–43.

Anderson, R. M., May, R. M., 1988. Epidemiological parameters of HIV transmission. Nature 333(6173), 514–9.

Anderson, R. M., May, R. M., 1991. Infectious diseases of humans: dynamics and control. Oxford University Press.

CDC, 2000. Tracking the Hidden Epidemics: Trends in STDs in the United States. Centers for Disease Control and Prevention.

Garnett, G. P., Anderson, R. M., 1993. Factors controlling the spread of HIV in heterosexual communities in developing countries: patterns of mixing between different age and sexual activity classes. Philos Trans R Soc Lond B Biol Sci 342(1300), 137–59.

Garnett, G. P., Anderson, R. M., 1994. Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations. IMA J Math Appl Med Biol 11(3), 161–92.

Grassly, N. C., Garnett, G. P., 2005. The future of the HIV pandemic. Bull World Health Organ 83(5), 378–82.

Gregson, S., Nyamukapa, C. A., Garnett, G. P., Mason, P. R., Zhuwau, T., Caraël, M., Chandiwana, S. K., Anderson, R. M., 2002. Sexual mixing patterns and sex-differentials in teenage exposure to HIV infection in rural Zimbabwe. Lancet 359(9321), 1896–903.

Haavio-Mannila, E., Kontula, O., Kuusi, E., 2001. Trends in Sexual life, measured by national sex surveys in Finland in 1971, 1992 and 1999, and a comparison to a sex survey in St. Petersburg in 1996. Helsinki: The Population Research Institute, The Family Federation of Finland.

Jæger, A. B., Gramkow, A., Sørensen, P., Melbye, M., Adami, H. O., Glimelius, B., Frisch, M., 2000. Correlates of heterosexual behavior among 23-87 year olds in Denmark and Sweden, 1992-1998. Arch Sex Behav 29(1), 91–106.

Johnson, A. M., Mercer, C. H., Erens, B., Copas, A. J., McManus, S., Wellings, K., Fenton, K. A., Korovessis, C., Macdowall, W., Nanchahal, K., Purdon, S., Field, J., 2001. Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours. Lancet 358(9296), 1835–42.

Johnson, A. M., Wadsworth, J., Wellings, K., Bradshaw, S., Field, J., 1992. Sexual lifestyles and HIV risk. Nature 360(6403), 410–2.

Korenromp, E. L., Van Vliet, C., Grosskurth, H., Gavyole, A., Van der Ploeg, C. P., Fransen, L., Hayes, R. J., Habbema, J. D., 2000. Model-based evaluation of single-round mass treatment of sexually transmitted diseases for HIV control in a rural African population. AIDS 14(5), 573–93.

Kretzschmar, M., van Duynhoven, Y. T., Severijnen, A. J., 1996. Modeling prevention strategies for gonorrhea and Chlamydia using stochastic network simulations. Am J Epidemiol 144(3), 306–17.

Lewin, B., Fugl-Meyer, K., Helmius, G., Lalos, A., Månson, S., 1998. Sex in Sweden 1996. National Institute of Public Health.

McQuillan, G. M., Coleman, P. J., Kruszon-Moran, D., Moyer, L. A., Lambert, S. B., Margolis, H. S., 1999. Prevalence of hepatitis B virus infection in the United States: the National Health and Nutrition Examination Surveys, 1976 through 1994. Am J Public Health 89(1), 14–8.

Melbye, M., Biggar, R. J., 1992. Interactions between persons at risk for AIDS and the general population in Denmark. Am J Epidemiol 135(6), 593–602.

Mertz, K. J., McQuillan, G. M., Levine, W. C., Candal, D. H., Bullard, J. C., Johnson, R. E., St Louis, M. E., Black, C. M., 1998. A pilot study of the prevalence of chlamydial infection in a national household survey. Sex Transm Dis 25(5), 225–8.

Morris, M., Kretzschmar, M., 1997. Concurrent partnerships and the spread of HIV. AIDS 11(5), 641–8.

Nakashima, A. K., Rolfs, R. T., Flock, M. L., Kilmarx, P., Greenspan, J. R., 1996. Epidemiology of syphilis in the United States, 1941–1993. Sex Transm Dis 23(1), 16–23.

Nnko, S., Boerma, J. T., Urassa, M., Mwaluko, G., Zaba, B., 2004. Secretive females or swaggering males? An assessment of the quality of sexual partnership reporting in rural Tanzania. Soc Sci Med 59(2), 299–310.

Stover, J., Walker, N., Garnett, G. P., Salomon, J. A., Stanecki, K. A., Ghys, P. D., Grassly, N. C., Anderson, R. M., Schwartländer, B., 2002. Can we reverse the HIV/AIDS pandemic with an expanded response? Lancet 360(9326), 73–7.

Thrall, P. H., Antonovics, J., Dobson, A. P., 2000. Sexually transmitted diseases in polygynous mating systems: prevalence and impact on reproductive success. Proc Biol Sci 267(1452), 1555–63.

Wang, C., Hawes, S. E., Gaye, A., Sow, P. S., Ndoye, I., Manhart, L. E., Wald, A., Critchlow, C. W., Kiviat, N. B., 2007. HIV prevalence, previous HIV testing, and condom use with clients and regular partners among Senegalese commercial sex workers. Sex Transm Infect 83(7), 534–40.

Ward, H., Day, S., Green, A., Cooper, K., Weber, J., 2004. Declining prevalence of STI in the London sex industry, 1985 to 2002. Sex Transm Infect 80(5), 374–6.

Ward, H., Day, S., Weber, J., 1999. Risky business: health and safety in the sex industry over a 9 year period. Sex Transm Infect 75(5), 340–3.

Wiederman, M. W., 1997. The truth must be in here somewhere: Examining the gender discrepancy in self reported lifetime number of sex partners. Journal of Sex Research 34, 375–386.

Figure legends

Figure 1: Disease prevalence eventually reaches an equilibrium value. Left panel: sex partner distributions for the two genders (identical in this simulation). Right panel: the resulting change in disease prevalences over time when individuals are allowed to interact according to the rules of our model. Each unit on the x-axis corresponds to 20 time-steps in the simulation. Note that after a sufficient number of time-steps have passed, disease prevalence will reach an equilibrium value. When both genders in the simulation have the same sex partner distribution, the size of this equilibrium only depends on p_{inf} and N_p.

Figure 2: Equilibrium prevalence as a function of $N_p p_{inf}$. The relationship between equilibrium prevalence, p_{inf}, and N_p was investigated empirically by performing simulations for a wide range of different p_{inf} and N_p values. Equilibrium prevalences for each set of parameters was averaged over the last 400 time-steps in the simulation (after confirming prior convergence) and were furthermore averaged over 10 independent runs. Several different distribution types were investigated (normal, mixture of two normals, Gumbel, discrete). The x-axis shows the value of $R_0 = N_p p_{inf}$ while the y-axis shows the equilibrium prevalence, f_I. The data points are the equilibrium values found by simulation, while the curve shows the theoretically predicted prevalence found using equation 4.

Figure 3: Sex partner distributions with different variance result in different equilibrium prevalence. Left panels: sex partner distributions for the two genders. Right panels: the corresponding development of disease prevalence over time. Each unit on the x-axis corresponds to 20 time-steps in the simulation. In the top three rows the male variance is larger than the female variance (with the same ratio between male and female variance in all three cases). In the bottom row the two genders have the same variance but different kurtosis. Top row: both genders have normal sex partner distributions (and therefore have the same kurtosis and skew). Second row: female distribution normal, male distribution a mixture of two normals. Third row: female distribution normal, male distribution Gumbel. Bottom row: males and females have discrete distributions tailored to give the same variance but different kurtosis. For each set of distributions 10 independent runs were performed and the prevalences at each time-step averaged.

Figure 4: Difference in gender prevalences is determined by N_p, p_{inf}, and the variance ratio. This plot shows the empirically determined equilibrium prevalence of females (black crosses) and males (colored symbols) as a function of the ratio between the variance of the male and female sex partner distributions. A variance ratio of one means the sex partner distributions for the two genders have the same variance, while ratios larger than one means the male sex partner distribution is wider than the female distribution. Equilibrium prevalence values (y-axis) were determined by averaging over the results from ten independent simulation experiments (for the same values of N_p, p_{inf}, and variance ratio). Data point shapes (triangles, circles, and squares) indicate the male sex partner distribution type (mixture of normals, Discrete, and Gumbel respectively), while colors indicate the values of the N_p and p_{inf} parameters. The dotted line shows the equilibrium prevalence predicted using equation 4 (this is the same for all experiments summarized in this plot since they all have $N_p p_{inf} = 8$). Note that female equilibrium prevalences are identical to the predicted prevalence regardless of distribution type and other parameter values. Also note that male prevalence depends on the variance ratio in a manner determined by N_p and p_{inf} but not by distribution type.

Figure 5: Prediction of male equilibrium prevalence when gender sex partner distributions differ. This plot shows the relationship between the male prevalence predicted from N_p, p_{inf}, and the variance ratio using equation 5 (x-axis) and the empirically observed male equilibrium prevalence (y-axis). A total of 60 different simulations were performed covering three different distribution types (discrete, mixture of two normals, and Gumbel), and a wide range of different values of N_p, p_{inf} and the variance ratio. The N_p and p_{inf} values were chosen to cover two different values of the product $N_p p_{inf}$. Each experiment was run 10 times and the equilibrium prevalences averaged over the runs. The correlation between predicted and observed values is very good with Pearson $r^2 = 0.99$, although there is a small, apparently systematic, error that gets worse for lower male prevalence values.

Table 1: Summary of published data on dispersion over the number of life-time heterosexual partners for men and women. "Study": Country and year of study. "Variance": variance of sex partner distribution.

"IQR": Interquartile range of sex partner distribution. "Ratio": ratio between male and female variances or interquartile ranges. Footnotes indicate when further information about distinct age groups is available in the publication (but note that variance and IQR values reported in the table have been computed across all age groups).

Table 1

Study	Variance (mean)		IQR (median)		Ratio
	Male	Female	Male	Female	
Britain 1992[a]	6575 (9.9)	165 (3.4)	-	-	39.8
Britain 2000[b]	1239 (12.7)	94 (6.5)	-	-	13.2
Denmark 1992[c]	-	-	10 (5.2)[g]	6 (3.8)[g]	1.7
Sweden 1996[d]	1173 (14.5)	70 (6.7)	-	-	16.8
Den/Swe 2000[e]	-	-	12 (7.7)[g]	3 (2.7)[g]	4.0
Finland 1971[f]	353 (11.0)	16 (2.6)	-	-	22.1
Finland 1992[f]	369 (13.8)	67 (5.2)	-	-	5.5
Finland 1999[f]	420 (14.4)	88 (6.6)	-	-	4.8
Russia 1996[f]	310 (11.5)	31 (4.2)	-	-	10.0

[a](Johnson et al., 1992); age groups: 16-24 / 25-34 / 35-44 / 45-59
[b](Johnson et al., 2001); age groups: 16-24 / 25-34 / 35-44
[c](Melbye and Biggar, 1992); age groups: 18-19 / 20-55 in 5 year bins / 55+
[d](Lewin et al., 1998); age groups: 18-73, all data points kindly made available
[e](Jæger et al., 2000); age groups: By decade of birth
[f](Haavio-Mannila et al., 2001); age groups: 18-34 / 35-54 / 55-74
[g]Values computed by measuring bar lengths on plots in publication